Journal of the Operations Research
Society of Japan
Vol. 40, No. 2, June 1997

最小絶対値法による回帰分析

末吉俊幸
東京理科大学

(受理 1996 年 2 月 15 日；再受理 1996 年 4 月 16 日)

和文概要　本研究の目的は最小絶対値法が現在広く使われている最小二乗法に匹敵しうる程実用性の高い回帰分析手法であることを示すことにある．本論文の前半では，18 世紀にさかのぼり，その歴史的考察を行うとともに，最小絶対値法を目標計画法の視点で考察する．後半では，最小絶対値法に関する統計理論とその統計的検定への応用を示した．重要なことは，最小絶対値法も最小二乗法も数理計画法でモデル化され，従来とは違った理論展開と応用可能性が開けることにある．

1. はじめに

本研究では最小絶対値法 (Least Absolute Value Estimation) による回帰分析を歴史的，理論的視点から考察する．この推定手法は目標計画法 (Goal Programming) でモデル化され，LP (Linear Programming) で解けることから OR (Operations Research) とも深い関係を持つ．OR の文献として最初にこの最小絶対値法による回帰分析を提唱したのはチャーンズとクーパー [9] で，その論文の発表以来，数多くの研究がこの分野でなされてきた．例えば，アルゴリズムの開発 [1, 4, 7, 30]，その解の理論研究 [6, 25, 26, 42]，シミュレーションによって最小二乗法との比較研究 [14, 28, 31]，さらに様々な分野への応用 [10, 35, 37, 38, 39, 40] などがある．[この分野の研究を集大成した本として [5, 19] があるので参考にされたい．]

本研究の目的は最小絶対値法が伝統的に使われる最小二乗法に匹敵しうるほど実用性の高い回帰分析手法であることを伝えることにあるが，もちろん，現在使われている最小二乗法を否定するつもりはなく，むしろ伝統的な手法の価値を認めつつ，目標計画法という OR 手法で，回帰分析がモデル化され，従来の最小二乗法とは別な形の回帰分析手法が存在することを示すことにある．

本研究の構成は，まず 18 世紀にさかのぼり，最小二乗法と最小絶対値法を作りだした人々の理論背景を考察する．次に，その科学史の流れに沿って，それら二つの回帰分析法の関係を論じる．さらに，最小絶対値法による統計的検定を具体例を使いながら考察する．最後に，本論文のまとめと将来の研究課題を論じる．

2. 最小絶対値法と最小二乗法の歴史

統計学の歴史に関する文献 (e.g., [21, 23, 34]) によると，16 世紀後半に芽生えつつあった回帰分析手法は 18 世紀になり Roger J. Boscovich (1711–1787) によって研究手法として体系化された．彼の考えを説明する為に，独立変数 (x) と従属変数 (y) で表わされたデータに対して $y=\beta_0+\beta_1 x$ の形をした回帰式を当てはめたと考える．これらの回帰係数 (β_0 と

β_1) を推定する回帰規準として，Boscovich は

$$\mathrm{MIN}\sum_{i=1}^{n}|y_i-(\beta_0+\beta_1x_i)| \tag{1}$$

を提唱した．この式では誤差の絶対値の合計を最小化している．添字 “i” はデータの測定順序を示し，“n” はデータのサンプル数を示している．Boscovich はこの回帰規準を使い，1757 年にローマ近郊の子午線を実測している [23]．この (1) 式は最小絶対値法の原形であると考えて良い．

次に，この回帰分析規準 (1) を最良のものと考え，そのアルゴリズムを最初に提案したのは Laplace [Piere Simmon Marquis de Laplace, 1749–1827] である．Laplace のアルゴリズムを説明するために，$\beta_0=0$ と $x_i>0$, $(i=1,\cdots,n)$ を仮定する．さらに，データを $y_1/x_1\geq y_2/x_2\geq\cdots\geq y_n/x_n$ となるように再配列する．ここで，もし

$$\begin{aligned} x_1+x_2+\cdots+x_{r-1} &< x_r+x_{r+1}+\cdots+x_n \quad \text{と} \\ x_1+x_2+\cdots+x_r &> x_{r+1}+x_{r+2}+\cdots+x_n \end{aligned}$$

を満たす整数 r が存在する時，β_1 の推定値は $\hat{\beta}_1=y_r/x_r$ で決められる．

さて，[34] の研究によると 1795 年頃まで Laplace は最小絶対値法を使っていたが，それ以降この回帰規準にあまり注目しなくなっている．その理由は彼の考えだしたアルゴリズムでは小さな限られた問題しか解くことができず，現実の応用が極めて難しい所にあった．この回帰分析の計算に関する問題を最初に解いたのは Gauss (Carl Friedrich Gauss, 1775–1855) であった．1795 年，Gauss が 20 才の時に，(1) にかわって最小二乗法による回帰分析規準を考えだし，測量のデータ分析に使いはじめていた．この回帰分析規準は

$$\mathrm{MIN}\sum_{i=1}^{n}[y_i-(\beta_0+\beta_1x_i)]^2 \tag{2}$$

で表現される．この規準の最も重要な特徴は微分可能性にある．(2) を微分することで，かなり大きなサンプルをもつ回帰係数でも推定することが可能になったわけである．面白いことに，Gauss にとってこの最小二乗法はあまりにも自明なことで，誰でもが，使っているものと思いこみ，その発表を 1821 年までのばしている [21]．

最初にこの最小二乗法を紹介したのは Legendre [Adrien Marie Legendre, 1752–1833] で，1805 年に “彗星の軌道を決める為の手法” として発表している．彼の貢献は現在広く知られている “正規方程式” を作りだしたことにある [23]．

Gauss は最小二乗法の発見後，1797 年から 1798 年の二年間，誤差の確率分布の研究に専念し，正規分布を発見している．さらに，もし誤差が正規分布に従うならば，最小二乗法によって推定された回帰係数が最尤推定値に一致することを証明している．[この最尤法は Daniel Bernoulli (1700–1782) によってすでに研究されていた．] 不幸にして，この Gauss の研究成果も 1809 年までその発表をのばされている [21]．

さて，1809 年に Gauss が研究成果を発表した後すぐに，Laplace は “中心極限定理” をまとめている [34]．[中心極限定理についての記述は [41] を参照されたい．] この中心極限定理と最尤法によって最小二乗法の基礎が確立され，今日に至るまで様々な分野で応用されていることは周知のことである．

本研究では微分不可能という理由で科学史の中から消えた最小絶対値法に焦点をあて，目標計画法を使うと，どのような回帰分析になるかを考察してみる．

3. 目標計画法と最小絶対値法

Boscovich の時代から解けなかった最小絶対値法による回帰分析を可能にしたのは目標計画法とコンピュータの発達による．ここでは具体的にどのように最小絶対値法を目標計画法でモデル化するかを説明するために，(1) を一般化した重回帰分析として問題を取り扱う．したがって，i 番目のデータに関する従属変数 (y_i) と m 個の独立変数ベクトル $\{X_i = (1, x_{i1}, \cdots, x_{im})\}$ の関係は $y_i = X_i\beta$ で表現される．ここで，β は回帰係数の列ベクトルで，$\beta = (\beta_0, \beta_1, \cdots, \beta_m)^{\mathrm{T}}$ で表現される．このβ は $m+1$ の要素を持つ．上つき添字 "T" はベクトルの転置を示している．

この重回帰分析のための最小絶対値法は

$$\mathrm{MIN}\sum_{i=1}^{n}|y_i - X_i\beta| \tag{3}$$

としてモデル化される．この (3) を目標計画法で表現するために，次の正と負の誤差に関する変数を導入する:

$$\delta_i^+ = 1/2\{|y_i - X_i\beta| + (y_i - X_i\beta)\}, \tag{4-1}$$

$$\delta_i^- = 1/2\{|y_i - X_i\beta| - (y_i - X_i\beta)\}. \tag{4-2}$$

ここで，δ_i^+ と δ_i^- はそれぞれ誤差の正と負の部分を表わし，$\delta_i^+ > 0, \delta_i^- > 0$ と $\delta_i^+ \cdot \delta_i^- = 0$ を常に満たす必要がある．この非線形の条件 $(\delta_i^+ \cdot \delta_i^- = 0)$ は $\delta_i^+ > 0$ と $\delta_i^- > 0$ が同時に起こるのを防ぐために必要である [8].

次に，(4-1) から (4-2) を引くと，

$$\delta_i^+ - \delta_i^- = y_i - X_i\beta, \tag{5-1}$$

また，(4-1) と (4-2) を足すと，

$$\delta_i^+ + \delta_i^- = |y_i - X_i\beta|, \tag{5-2}$$

が得られる．

(5-1) と (5-2) より最小絶対値法 (3) は次のような目標計画法で定式化される．

$$\begin{array}{ll} \text{目的関数} & \mathrm{MIN}\sum_{i=1}^{n}(\delta_i^+ + \delta_i^-) \\ \text{制約} & X_i\beta + \delta_i^+ - \delta_i^- = y_i, \\ & \delta_i^{+*} \geq 0, \quad \delta_i^{-*} \geq 0, \quad i = 1, \cdots, n. \end{array} \tag{6}$$

その双対問題は

$$\begin{array}{ll} \text{目的関数} & \mathrm{MAX}\sum_{i=1}^{n} w_i y_i \\ \text{制約} & \sum_{i=1}^{n} w_i x_{ij} = 0, \quad j = 1, \cdots, m \\ & -1 \leq w_i \leq 1, \quad i = 1, \cdots, n \end{array} \tag{7}$$

となる．ここで，w_i^* は i 番目の双対変数であり，x_{ij} は j 番目の独立変数の i 番目の実測値である．この (6) 式と (7) 式は目的関数の重み付けがすべて "1" の目標計画法と考えて良い．したがって，LP のアルゴリズムで解くことができる．

この最小絶対値法は回帰分析手法として様々な性質を持つが，ここでは四つの重要な性質を述べておく．はじめに，その最適値において，

$$\sum_{i=1}^{n}(\delta_i^{+*}+\delta_i^{-*})=\sum_{i=1}^{n}w_i^* y_i$$

が成り立つ．ここで，上つき添字 "*" は最適値を示している．この w_i^* は y_i が1単位増加すると最小化された目的関数がどれだけ増加するかを示している．さらに，双対の相補性により次の関係が成り立つことが分かる．

(a) $w_i^*=1 \quad \Leftrightarrow \quad \delta_i^{+*}>0$ かつ $\delta_i^{-*}=0$
(b) $w_i^*=-1 \quad \Leftrightarrow \quad \delta_i^{+*}=0$ かつ $\delta_i^{-*}>0$
(c) $-1<w_i^*<1 \quad \Leftrightarrow \quad \delta_i^{+*}=\delta_i^{-*}=0$

このように，w_i^* を調べることによって推定された回帰式と誤差の関係が分かる [36].

二番目に，最小絶対値法はその最適解を無限個作りだす可能性がある．この特徴は "退化" と呼ばれ，LP のすべての問題に付随しているもので，ある i 番目のデータに対して最適化された δ_i^{+*} ならびに δ_i^{-*} がシンプレックス法の基底を形成し，その最適値が零になる時に起こる現象である．この退化を双対問題で言い換えると，i 番目のデータが推定された回帰式の上にあり，その双対変数が $w_i^*=1$ となる場合，又は，$w_i^*=-1$ の場合に起こる現象と考えてよい．この退化が起こった場合，最小絶対値法は無限個の回帰式をその解として作りだす欠点を持つ．[最小二乗法の場合，解の一意性は独立変数の一次独立性が成り立てば，常に保証されている．]

三番目に，(6) は様々な情報をアプリオリー（先験的条件）な形で組み入れることができる．例えば，経営者の勘，経験，理論などから $\hat{\beta}$ や $\hat{y}$ の上限と下限を設定し，それらを制約式の形で (6) の中に入れることができる．その結果，推定された答が分析者やその利用者にとってより受け入れやすいものになる．ただ，問題なのは，それらのアプリオリー（先験的条件）な情報が LP の制約式の形で表現される必要のある点である．[[11] の研究の中で (6) とアプリオリー (先験的条件) な情報を組み入れることで，Multicollinearity の問題を解決しているので参照されたい．]

次に，最小絶対値法と最小二乗法という二つの推定法によって得られた回帰係数の推定精度についてコメントする．この問題は研究者の間で長い議論の歴史を持つ．[例えば，Eddington (1914)[20, P.147] と Fisher (1912) [22, P.762]．] 要約すると，誤差が正規分布に従う場合，最小二乗法の方が最小絶対値法よりもより信頼度の高い係数推定値を与える．しかしながら，もし誤差が非正規分布（特に，Outlier と呼ばれるデータの外れ値を含む時）をする場合，係数推定値の信頼度において，それら二つの関係は逆転する．さらに，誤差が正規分布する時，最小二乗法による回帰係数は最尤推定値に一致するが，誤差が Laplace 分布をする時は最小絶対値法による回帰係数が最尤推定値となる [5].

4. 不偏推定と最小絶対値法

β を中心にその推定値 $\hat{\beta}$ が分布していることがのぞまれることは当然である．これを式で書き表わせば，$E(\hat{\beta})=\beta$ となる．本研究では，最小絶対値法を使い，どのように不偏推定値を求めるかを考察してみる [5, 32].

はじめに，この不偏推定値を得るために，誤差 (ε) が対称 (symmetric) に分布し，$E(\varepsilon)=0$ になると仮定する．

[定義] β とその推定値 $(\hat{\beta})$ が$\beta - \hat{\beta}(\varepsilon) = -[\beta - \hat{\beta}(-\varepsilon)]$ をすべての ε に対して満たす時 "antisymmetric (反対称)" と定義する．

例えば，最小二乗法による回帰係数 (β_{LS}) はこの反対称性を満たす．つまり，

$$\begin{aligned}
&\beta_{\mathrm{LS}} = (X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}Y,\\
&\beta_{\mathrm{LS}}(\varepsilon) = (X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}(X\beta+\varepsilon) = \beta + (X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}\varepsilon,\\
&\beta_{\mathrm{LS}}(-\varepsilon) = (X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}(X\beta-\varepsilon) = \beta - (X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}\varepsilon.
\end{aligned}$$

したがって，$\beta - \beta_{\mathrm{LS}}(\varepsilon) = -[\beta - \beta_{\mathrm{LS}}(-\varepsilon)]$ となり，反対称性の条件を満たす．

次に，(6) 式の最小絶対値法を書き直すと，

$$\begin{array}{ll}
\text{目的関数} & \mathrm{MIN}\displaystyle\sum_{i=1}^{n}\delta_i\\
\text{制約} & -\delta_i \le y_i - X_i\beta \le \delta_i, \quad i=1,\cdots,m\\
 & \delta_i \ge 0
\end{array} \tag{8}$$

となる．

さらに，(8) 式は行列で表現すると，

$$\begin{array}{ll}
\text{目的関数} & \mathrm{MIN}\delta^{\mathrm{T}}e\\
\text{制約} & -\delta \le Y - X\beta \le \delta\\
 & \delta \ge 0
\end{array} \tag{9}$$

となる．

ここで，回帰係数ベクトル (β) は正と負のどちらにでもなれるので，$\beta = \beta_{\mathrm{P}} - \beta_{\mathrm{N}}$ を $\beta_{\mathrm{P}} \ge 0$ と $\beta_{\mathrm{N}} \ge 0$ の 2 つのベクトルに分けると，(9) 式は

$$\begin{array}{ll}
\text{目的関数} & \mathrm{MIN}\delta^{\mathrm{T}}e\\
\text{制約} & \begin{bmatrix} -X & X & -\mathrm{I}\\ X & -X & -\mathrm{I}\end{bmatrix}\begin{bmatrix}\beta_{\mathrm{N}}\\ \beta_{\mathrm{P}}\\ \delta\end{bmatrix} \le \begin{pmatrix} -Y\\ Y\end{pmatrix}\\
 & \beta_{\mathrm{P}} \ge 0, \quad \beta_{\mathrm{N}} \ge 0, \quad \delta \ge 0
\end{array} \tag{10}$$

となる．さらに，この (10) 式は

$$\begin{array}{ll}
\text{目的関数} & \mathrm{MIN}\delta^{\mathrm{T}}e\\
\text{制約} & \begin{bmatrix} -X & X & -\mathrm{I}\\ X & -X & -\mathrm{I}\end{bmatrix}\begin{bmatrix}\beta_{\mathrm{N}}\\ \beta_{\mathrm{P}}\\ \delta\end{bmatrix} \le \begin{pmatrix} Y\\ -Y\end{pmatrix}\\
 & \beta_{\mathrm{P}} \ge 0, \quad \beta_{\mathrm{N}} \ge 0, \quad \delta \ge 0
\end{array} \tag{11}$$

と同等である．

次に，antisymmetric の推定値 $(\beta_{\mathrm{A}} = \beta_{\mathrm{A}}^{(\mathrm{P})} - \beta_{\mathrm{A}}^{(\mathrm{N})}, \beta_{\mathrm{A}}^{(\mathrm{P})} \ge 0, \beta_{\mathrm{A}}^{(\mathrm{N})} \ge 0)$ を使い，新しい変数ベクトル $B_{\mathrm{P}}\ (= \beta_{\mathrm{P}} + \beta_{\mathrm{A}}^{(\mathrm{N})})$ と $\beta_{\mathrm{N}}\ (= \beta_{\mathrm{N}} + \beta_{\mathrm{A}}^{(\mathrm{P})})$ を導入すると，(10) 式と (11) 式はそれ

ぞれ，

$$\begin{array}{ll} \text{目的関数} & \mathrm{MIN}\delta^{\mathrm{T}}e \\ \text{制約} & \begin{bmatrix} -X & X & -\mathrm{I} \\ X & -X & -\mathrm{I} \end{bmatrix} \begin{bmatrix} B_{\mathrm{N}} \\ B_{\mathrm{P}} \\ \delta \end{bmatrix} \leq \begin{pmatrix} -Y + X\beta_{\mathrm{A}} \\ Y - X\beta_{\mathrm{A}} \end{pmatrix} \\ & B_{\mathrm{P}} \geq 0, \quad B_{\mathrm{N}} \geq 0, \quad \delta \geq 0 \end{array} \tag{12}$$

と

$$\begin{array}{ll} \text{目的関数} & \mathrm{MIN}\delta^{\mathrm{T}}e \\ \text{制約} & \begin{bmatrix} -X & X & -\mathrm{I} \\ X & -X & -\mathrm{I} \end{bmatrix} \begin{bmatrix} B_{\mathrm{N}} \\ B_{\mathrm{P}} \\ \delta \end{bmatrix} \leq \begin{pmatrix} Y - X\beta_{\mathrm{A}} \\ -Y + X\beta_{\mathrm{A}} \end{pmatrix} \\ & B_{\mathrm{P}} \geq 0, \quad B_{\mathrm{N}} \geq 0, \quad \delta \geq 0 \end{array} \tag{13}$$

になる．

これら (12) 式と (13) 式を使い，不偏推定値は次のアルゴリズムで求めることができる [5, 32].

[アルゴリズム]

(a) (12) 式と (13) 式をそれぞれ確率 50% で選ぶ．

(b) 選ばれた式の最適値を $\hat{B}_{\mathrm{P}}, \hat{B}_{\mathrm{N}}$ とする．

(c) β の推定値を $\beta_{\mathrm{LAV}} = \hat{B}_{\mathrm{P}} - \hat{B}_{\mathrm{N}} + \beta_{\mathrm{A}}$ とする．

[結果] β_{LAV} は β の不偏推定値である．

重要なことは，β_{A} を見つけだすことで，最小絶対値法でも不偏推定量になりうることである．

5. 理論的背景

ここでは，これまでの結果をふまえて，最小絶対値法による統計的検定を説明する．その狙いは統計的検定を通じて，最小二乗法と最小絶対値法の違いを説明することにある．この目的を達成するために具体例を使って考察してみる．

2 章で説明したように，最小絶対値法に関する歴史上の問題点はその解を得るためのアルゴリズムであった．このアルゴリズムは目標計画法によってモデル化され，LP で解けることが明らかになった．アルゴリズムの発見後，研究者の間で関心になったテーマはどのように統計的検定を行うかということである．この領域での最初の論文は [6] と [25] の中に見られる．この理論研究の発表後，多くの論文 (e.g., [2, 3, 18, 24, 42]) が発表されている．

本章では，[6] の理論研究を説明するために，その理論研究の仮定を述べることから始める．

[仮定]

(a) $f(F^{-1}(1/2)) = f(0) > 0$: 誤差の確率分布はメディアン $f(0)$ において連続で，正の確率密度関数を持つ．

(b) $\lim_{n\to\infty} 1/n(X^{\mathrm{T}}X) = D$ において，D は正定値を持つ．

[結果] この二つの仮定が満たされると，最小絶対値法で推定される回帰係数$\hat{\beta}_n$ は

$$\sqrt{n}(\hat{\beta}_n - \beta) \xrightarrow{D} N(0, \lambda^2(X^{\mathrm{T}}X)^{-1})$$

を満たす．ここで，λ は $1/[2f(0)]$ である．

この結果が示すことは，推定された回帰係数はサンプル数を大きくすると，共分散行列 $\lambda^2(X^{\mathrm{T}}X)^{-1}$ をした正規分布に近づくことが漸近的に証明している．[詳しい証明は [3] と [6] の中にあるので参照されたい．]

さて，$f(0)$ はメディアンにおける誤差の確率密度関数を表わしている．したがって σ^2 を誤差の分散とすると，$\lambda^2 < \sigma^2$ が成り立つ場合，最小絶対値法の方が最小二乗法より小さい共分散行列を持つことが分かる．例えば，誤差が Laplace 分布に従うとすると，$f(0) = 1/(\sqrt{2}\sigma)$ となり，$\lambda^2 = \sigma^2/2$ が得られる．したがって，Laplace 分布において，最小絶対値法は最小二乗法よりも，半分の共分散行列を持つことが漸近的に示される．[Laplace 分布の確率密度関数は $f(x) = c/2 \exp^{-|x|/c}$ $(-\infty < x < \infty)$ で表される．この分布の形を決める定数は c で表している．この確率密度関数からサンプル平均の分散 $\mathrm{Var}(\bar{x})$ とサンプルメディアンの分散 $\mathrm{Var}(m)$ のそれぞれは $\mathrm{Var}(\bar{x}) = 1/[2c^2]$ と $\mathrm{Var}(m) = 1/[4c^2]$ で求められる．このことから $\mathrm{Var}(\bar{x}) = 1/[2c^2] = \sigma^2$ の場合，$\mathrm{Var}(m) = 1/[4c^2] = \lambda^2 = \sigma^2/2$ が得られる．さらに，$f(m) = f(0) = 1/(\sqrt{2}\sigma)$ が求められる．]

さて，最小絶対値法によって推定された回帰係数に関する asymptotics はλ が既知と仮定している．現実の問題では，λ が既知であることはあまりなく，何らかな形で，λ を推定する必要がある．このλ の推定値を得るために様々な研究 (e.g., [15, 16, 26, 27, 29, 33]) がなされてきた．例えば，[12, p.369], [13] の中で，サンプルが極めて大きい時，λ の推定値は $\hat{\lambda} = 1/[2\hat{f}(0)]$ で表現される．さらに，

$$\hat{\lambda} = 1/[2\hat{f}(0)] = [\hat{\varepsilon}_{(t)} - \hat{\varepsilon}_{(s)}]/[2(t-s)/n] \tag{14}$$

となることが, [13, p.40] と [14, p.846] の中で示されている．ここで，$\hat{\varepsilon}_{(t)}$ と$\hat{\varepsilon}_{(s)}$ はメディアンを中心に順序づけられた t 番目と s 番目の誤差を表している．

さて，サンプル数が n 個ある場合，この t と s の順序はメディアンを中心に対称に順序づけられて，

$$t = [n/2] + v \quad \text{と} \quad s = [n/2] - v \tag{15}$$

として選ばれる．ここで $[\cdot]$ はある数の整数比を示している．また，v はメディアンの誤差からある幅をもたせるための整数値である．この$\hat{\lambda}$を求める上での注意点として，(a) $\hat{\varepsilon}_{(t)} > 0$, $\hat{\varepsilon}_{(s)} > 0$，また，(b) v はできるだけ小さい方が良いことなどが上げられる．[ここでの議論の詳しい証明は [17] や [29] の中で記述されているので参照されたい．]

ただ，主観的見解だが，$\hat{\lambda}$ の推定精度はサンプル数に依存する．Asymptotics という名前が示すように，(14) 式が成り立つのはかなり大きなサンプルを必要とする．問題なのは，サンプル数が小さくなった場合，違った v の値は違った $\hat{\lambda}$ をうみだし，推定精度が悪くなる可能性が強い．サンプル数が小さい場合，最良の v 値をどのように設定したら良いかが課題として残る．従来の研究を見ると [e.g., [17, 26, 27]]，この $\hat{\lambda}$ の推定の難しさから，$\hat{\lambda}$ なしに統計的検定を行う方法が考えだされたり，$\hat{\lambda}$ の推定精度を良くしようとする試みがなされている [e.g., [17, 27]]．本研究では，最小二乗法との関係で記述するために，$\hat{\lambda}$ を推定する方法で解説を進める．

6. 信頼限界の推定

この最小絶対値法に関する理論研究より，β_j の信頼限界は，

$$\hat{\beta}_j \pm Z_{\alpha/2}\hat{\lambda}(X^{\mathrm{T}}X)_{jj}^{-1/2} \tag{16}$$

で表現される．ここで，$(X^{\mathrm{T}}X)_{jj}^{-1/2}$ はその行列の (j,j) 要素の平方根を示している．また，添字 “α” は $100(1-\alpha)\%$ の信頼度を，また $Z_{\alpha/2}$ は標準正規分布から得られた確率を，それぞれ表わしている．

次に，ある独立変数ベクトル $X_0=(x_{01},x_{02},\cdots,x_{0m})^{\mathrm{T}}$ に対する予測値の平均値の $100(1-\alpha)\%$ の信頼限界は，

$$X_0\hat{\beta}\pm Z_{\alpha/2}\hat{\lambda}[X_0^{\mathrm{T}}(X^{\mathrm{T}}X)^{-1}X_0]^{1/2} \tag{17}$$

で求められる．

以上は回帰式の存在範囲を予想したのであるが，もし将来行われるであろう実験，つまり X_0 を指定して実験を行なったとして，その一つの実験に対する予測値の $100(1-\alpha)\%$ の信頼区間は誤差を加味して，

$$X_0\hat{\beta}\pm Z_{\alpha/2}\hat{\lambda}[1+X_0^{\mathrm{T}}(X^{\mathrm{T}}X)^{-1}X_0]^{1/2} \tag{18}$$

のように修正する必要がある．

7. 回帰係数の検定

帰無仮説 H_0: $\beta_1=\beta_2=\cdots=\beta_m=0$，対立仮説 H_1: ある j において $\beta_j\neq 0$ を検定するために

$$\hat{\beta}'^{\mathrm{T}}X'^{\mathrm{T}}X'\hat{\beta}'/\hat{\lambda}^2 \tag{19}$$

を利用して自由度が m の χ^2 分布で検定する．ここで，添字 “$'$” は切片を示す β_0 を取り除く回帰係数ベクトルを表わしている．ここでの狙いは従属変数 (y) と独立変数 (x) の部分集合が線形関係を持つかどうかを確かめることにある．

次に，$H_0:\beta_j=0$，　$H_1:\beta_j\neq 0$ を検定する時は

$$\hat{\beta}_j^2/\hat{\lambda}^2c_{jj} \tag{20}$$

を使い，自由度 1 の χ 分布で検定する．ここで，c_{jj} は $(X^{\mathrm{T}}X)^{-1}$ の (i,j) 番目の対角要素を示している．

最後に，回帰係数の検定に次いて，1つのコメントをする．(20) 式は，[26] の研究によると，Wald 法とよばれ，この検定法の他に，2種の別の形の検定法があることが知られている．この別のアプローチは $\hat{\lambda}$ を必要とせずに，検定を行なえる長所を持っている．[詳しい説明は [17] の中にあるので参照されたい．]

8. 具体例による比較，検討

具体的なデータを使い，最小二乗法と最小絶対値法を比較，考察してみる．ここで使われているデータは表 1 に示され，従属変数 (y) と独立変数 (x_1,x_2) から成り立っている．

8.1 最小二乗法による分析結果

このデータから

$$(X^{\mathrm{T}}X)^{-1}=\begin{bmatrix} 0.214653 & -0.007491 & -0.000340 \\ -0.007491 & 0.001671 & -0.000019 \\ -0.000340 & -0.000019 & 0.000002 \end{bmatrix}$$

表 1: データ

データ番号	y	x_1	x_2
1	9.95	2	50
2	24.45	8	110
3	31.75	11	120
4	35.00	10	550
5	25.02	8	295
6	16.86	4	200
7	14.38	2	375
8	9.60	2	52
9	24.35	9	100
10	27.50(10000)	8	300
11	17.08	4	412
12	37.00	11	400
13	41.95	12	500
14	11.66	2	360
15	21.65	4	205
16	17.89	4	400
17	19.00	20	600
18	10.30	1	585
19	34.92	10	540
20	46.59	15	250
21	48.88	15	290
22	54.12	16	510
23	56.63	17	590
24	22.13	6	100
25	21.15	5	400

が得られ，最小二乗法による回帰式は

$$\begin{array}{rcccccc} y = & 2.279 & + & 2.793x_1 & + & 0.012x_2 \\ & (1.044) & & (0.092) & & (0.003) \end{array}$$

となる．この回帰係数の推定値とその精度は次のようにまとめられる．

変数名	係数推定値	標準偏差	t 値	prob $> t$
定数	2.279	1.044	2.183	0.040
x_1	2.793	0.092	30.323	0.000
x_2	0.012	0.003	4.269	0.000

さらに，分散分析は

要因	平方和	自由度	分散	分散比
回帰	6136.28	2	3068.14	604.37
残差	111.69	22	5.08	
総	6247.97	24		

となる．

この結果から各 β_j の $100(1-\alpha)\%$ の信頼区間は$\hat{\beta}_j \pm t_{1/2}\sqrt{\hat{\sigma}^2 c_{jj}}$ で求めることができる．例えば，β_1 の 95% 信頼区間は $2.55 \leq \beta_1 \leq 2.94$ として推定される．次に，分散分析表により $F_o = 604.37 > F_{5\%,22} = 3.44$ が得られ，帰無仮説 $H_0 : \beta_1 = \beta_2 = 0$ が 5% 有意水準で棄却される．

また，回帰係数の推定値とその精度に示されたように，各帰無仮説 $H_0 : \beta_1 = 0$ に関する検定をするために $t_0 = \hat{\beta}_j / \sqrt{\hat{\sigma}^2 c_{jj}}$ を利用する．例えば $H_0 : \beta_2 = 0$ に関する検定は $t_0 = 4.269$ が得られ，$t_{25\%,22} = 2.074$ より，この帰無仮説を棄却する．

8.2 最小絶対値法による分析結果

最小絶対値法によって推定された回帰式は

$$y = 3.6670 + 2.7912x_1 + 0.0066x_2$$

となる．

ここで $v = 3$ を (14) 式に導入して，$\hat{\lambda} = (0.4282 - (-0.1913))/12 = 1.2906$ を得る．[$v = 2$ の場合，$\hat{\lambda} = 1.1369$, 一方 $v = 4$ の場合，$\hat{\lambda} = 1.488$ となり，v の選び方によって $\hat{\lambda}$ の推定値が多少変化することが確かめられる．] ここでは，推定された $\hat{\lambda}^2$ と最小二乗法で得られた $\hat{\sigma}^2$ を比較すると，$\hat{\lambda}^2 = 1.66 < \hat{\sigma}^2 = 5.23$ となり最小絶対値法の方がより小さい共分散行列を持つことがこのデータでは確かめられる．

次に，最小二乗法によって得られた結果と比較する．はじめに，β_1 の 95% 信頼区間は $2.7912 \pm 1.96(1.29)\sqrt{0.001671}$，$2.688 \leq \beta_1 \leq 2.895$ となる．また帰無仮説 $H_0 : \beta_1 = \beta_2 = 0$ を検定するためのχ^2 分析は

χ^2 検定	自由度	有意水準
$\hat{\beta}'^{\mathrm{T}} X'^{\mathrm{T}} X' \hat{\beta}' / \hat{\lambda}^2 = 130062$	2	$\Pr(\chi^2 > 13006) << 0.01$

のようにまとめられ，明らかに $H_0 : \beta_1 = \beta_2 = 0$ は有意水準 1% で棄却される．さらに，帰無仮説 $H_0 : \beta_2 = 0$ を検定する χ^2 分析は

χ^2 検定	自由度	有意水準
$\hat{\beta}_2^2 / \hat{\lambda}^2 c_{22} = 22.47$	1	$\Pr(\chi^2 > 22.47) << 0.01$

としてまとめられ，$H_0 : \beta_2 = 0$ は有意水準 1% で棄却される．

8.3 異常値が回帰係数に与える影響

この表 1 のデータに異常値を入れ，それが二種類の回帰係数に与える影響を調べてみる．異常値として 10 番目のデータの y を 27.50 から 10000 にかえてみる．

表 2 は異常値がある場合とない場合に回帰係数がどのように変化するかをまとめてある．この表 2 からわかるように，最小絶対値法によって推定された回帰係数は異常値の存在にあまり影響されていない．ただ標準偏差にはある程度の影響がみられる．例えば，$\hat{\beta}_1$ は異常値がない場合に 2.791 と推定されるが，それがある場合でも 2.730 と推定されている．他の回帰係数にも同じような異常値に対する頑健性がみられる．これに対して，最小二乗法に

表 2: 異常値が回帰係数に与える影響

分析法 回帰係数	異常値がない場合		異常値がある場合	
	最小二乗法	最小絶対値法	最小二乗法	最小絶対値法
定数	2.279 (1.044)	3.667 (1.531)	533.888 (964.589)	4.136 (196.554)
$x1$	2.793 (0.092)	2.791 (0.103)	4.057 (85.101)	2.730 (13.245)
$x2$	0.012 (0.003)	0.007 (0.004)	-0.426 (2.546)	0.007 (0.492)

注) ( ) の中の数字は標準偏差を表している.

よって推定された回帰係数は異常値の存在にかなりの影響を受けていることが分かる. 例えば, 異常値がない場合に $\hat{\beta}_1 = 2.793$ と推定されるが, 異常値がある場合は$\hat{\beta}_1 = 4.057$ となる. $\hat{\beta}_0$ は 2.279 から 533.888 に変化し, $\hat{\beta}_2$ は 0.012 から -0.426 に変化している. [$\hat{\beta}_2$ の場合大きさだけでなくサインも正から負へ変化している.] このように表 2 から最小絶対値法の異常値に対する頑健性 (Robustness) を確かめることができる.

9. 結論と将来展望

本研究の目的は最小絶対値法が現在広く使われている最小二乗法に匹敵しうる程実用性の高い回帰分析手法であることを示すことにある. 本論文の前半では, 18 世紀にさかのぼり, その歴史的考察を行うとともに, 最小絶対値法を目標計画法の視点で考察した. 後半では, 最小絶対値法に関する統計理論とその統計的検定への応用を示した.

重要なことは, 最小絶対値法も最小二乗法も数理計画法でモデル化され, 解くこともでき, 数理計画法の視点でとらえると, 従来とは違った理論と応用が可能であることにある.

将来の研究テーマとして, サンプル数が有限の場合, 最小絶対値法によって推定された回帰式と誤差の分布 (e.g., 二項分布, Poisson 分布) の関係をどのように把握したら良いのか? また, データの異常値 (Outlier) にも, いろいろな種類があり, Leverage Point と呼ばれる異常値が存在する場合, 最小絶対値法の頑健 (Robustness) は非常に悪くなるので, この問題にどのように対応すべきか? さらに, ブートストラップ法 [36] とよばれるシミュレーション手法と最小絶対値法を組み合わせて, 回帰係数の分散を推定すると, 漸近的に求められた誤差の分布に関する仮定を必要としなくなるので, その組み合わせに基づく統計的検定をどのようにすべきか? 最後に, 最小絶対値法による回帰分析はマネジメント, 経済, 政策科学などの様々の分野で応用することができ, その応用の仕方をそれぞれの問題に合わせて考察する必要がある. 本研究はそれら諸問題を将来の研究課題とする.

参考文献

[1] Abdelmalek, N., "Efficient Methods for the Discrete Linear Approximation Problem", *Mathematics of Computation*, 29 (1975), pp.844–855.

[2] Amemiya, T., "Two Stage Least Absolute Deviations Estimators", *Econometrica*, 50 (1982), pp.689–711.

[3] Amemiya, T., *Advanced Econometrics*, Harvard University Press, Massachusetts (1985).

[4] Armstrong, R.D., E.L. Frome and D.S. Kung, "A Revised Simplex Algorithm for the Absolute Deviation Curve Fitting Problem", *Communication in Statistics-Simulation and Computation*, B8(2) (1979), pp.175–190.

[5] Arthanari, T.S. and Y. Dodge, "Mathematical Programming in Statistics", John Wiley & Sons, New York (1981)

[6] Bassett, G. and R. Koenker, "Asymptotic Theory of Least Absolute Error Regression", *Journal of the American Statistical Association*, 73 (1978), pp.618–622.

[7] Barrodale, I. and F. Roberts, "An Improved Algorithm for Discrete L1 Linear Approximation with Linear Constraints", *SIAM Journal of Numerical Analysis*, 15 (1978), pp.603–611.

[8] Charnes, A. and W.W. Cooper, "Goal Programming and Multiple Objective Optimization", *European Journal of Operational Research*, 1 (1977), pp.39–54.

[9] Charnes, A., W.W. Cooper and R.O. Ferguson, "Optimal Estimationof Executive Compensation by Linear Programming", *Management Science*, 1 (1955), pp.138–151.

[10] Charnes, A., W.W. Cooper and T. Sueyoshi, "Goal Programming/Constrained Regression Review of the Bell System Breakup", *Management Science*, 34 (1988), pp.1–26.

[11] Charnes, A., W.W. Cooper and T. Sueyoshi, "Least Squares/Ridge Regression and Goal Programming/Constrained Regression Alternatives", *European Journal of Operational Research*, 27 (1986), pp.146–157.

[12] Cramer, H., *Mathematical Methods of Statistics*, Princeton University Press, Princeton (1946).

[13] Dielman, T. and P. Pfaffenberger, "LAV (Least Absolute Value) Estimation in Linear Regression: A Review", *Studies in the Management Sciences* edited by S.H. Zanakis and J.S. Rustagi, 19 (1982), pp.31–52.

[14] Dielman, T. and R.C. Pfaffenberger, "Bootstrapping in Least Absolute Value Regression: An Application to Hypothesis Testing", *Communications in Statistics-Simulation and Computation*, 17 (1988), pp.843–856.

[15] Dielman, T. and R. C. Pfafffenberger, "Least Absolute Value Regression: Necessary Sample Sizes to Use Normal Theory Inference Procedures", *Decision Sciences*, 19 (1988), pp.734–743.

[16] Dielman, T. and P. Pfaffenberger, "Tests of Linear Hypotheses and LAV Estimation: A Monte Carlo Comparison", *Communications in Statistics-Simulation and Computation*, 19 (1990), pp.1179–1199.

[17] Dielman, T. and P. Pfaffenberger, "A Further Comparison of Tests of Hypotheses in LAV Regression", *Computational Statistics & Data Analysis*, 14 (1992), pp.375–384.

[18] Devroye, L. and L. Gyorfi, *Nonparametric Density Estimation: The L1 View*, John Wiley & Sons, New York (1992).

[19] Dodge, Y., *Statistical Data Analysis Based on the L1-Norm and Related Methods*, North-Holland, Amsterdam (1987).

[20] Eddington, A. S., *Stellar Movements and the Structure of the Universe*, Macmillan, London (1914).

[21] Eisenhart, C., "The Meaning of 'Least' in Least Squares", *Journal of the Washington*

Academy of Science, (1964), pp.24-33.

[22] Fisher, R.A., "A Mathematical Examination of the Methods of Determining the Accuracy of an Observation by the Mean Error and Mean Square Error", *Monthly Not. Roy. Astron. Soc.*, 80 (1920), pp.758–770.

[23] Harter, H.L., "The Method of Least Squares and Some Alternatives", *International Statistical Review*, 42 (1974), pp.147–174, 235–264 and 43 (1975), pp.1–44.

[24] Honore, B.E., "Trimmed LAD and Least Squares Estimation of Truncated and Censored Regression Models with Fixed Effects", *Econometrica*, 60 (1992), pp.533–565.

[25] Koenker, R. and G. Bassett, "Regression Quantiles", *Economerica*, 46 (1978), pp.33–50.

[26] Koenker, R. and G. Bassett, "Tests of Linear Hypotheses and L1 Estimation", *Economerica*, 50 (1982), pp.1577–1583.

[27] Koenker, R. and G. Bassett, "Robust Tests for Heteroscedasticity Based on Regression Quantiles", *Economerica*, 50 (1982), pp.43–61.

[28] Narula, S.C. and J.F. Wellington, "The Minimum Sum of Absolute Errors Regression: A State of the Art Survey", *International Statistical Review*, 50 (1982), pp.317–326.

[29] Powell, J.L., "Least Absolute Deviations Estimation for the Censored Regression Model", *Journal of Econometrics*, (1984).

[30] Rhee, W.T. and K.A. Rhee, "Minimization Technique for a Convex Function with Application to Multiple Regession Model", *Optimization*, 19 (1988), pp.253–267.

[31] Rosenberg, B. and D. Carlson, "A Simple Approximation of the Sampling Distribution of Least Absolute Residuals Regression Estimates", *Communications in Statistics-Simulation and Computation*, B6 (1977), pp.421–438.

[32] Sielken, R.L. and H.O. Hartley, "Two Linear Programming Algorithms for Unbiased Estimation of Linear Models", *Journal of the American Statistical Association*, 68 (1973), pp.639–641.

[33] Silverman, B.W., *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, London (1986).

[34] Stigler, S.M., "Studies in the History of Probability and Statistics. XXXII", *Biometrika*, 60 (1973), pp.439–445.

[35] Sueyoshi, T., "Estimation of Stochastic Frontier Cost Function Using Data Envelopment Analysis: An Application to the AT& T Divestiture", *Journal of the Operational Research Society*, 42 (1991), pp.463–477.

[36] Sueyoshi, T., "Empirical Regression Quantile", *Journal of the Operations Research Society of Japan*, 34 (1991), pp.250–262.

[37] Sueyoshi, T., "Divestiture of Nippon Telegraph & Telephone," *Management Science*, 42 (1996), pp.1326–1351.

[38] Sueyoshi, T., "Constrained Regression Median for Measuring Possible Salary Discrimination", *European Journal of Operational Research*, 77 (1994), pp.253–271.

[39] Sueyoshi, T., "Stochastic Frontier Production Analysis: Measuring Performance of Public Telecommunications in 24 OECD Countries", *European Journal of Operational Research*, 74 (1994), pp.466–478.

[40] Sueyoshi, T. and Y.L. Chang, "Goal Programming Approach for Regression Median", *Decision Sciences*, 20 (1989), pp.700–714.

[41] Theil, H., *Principles of Econometrics*, John Wiley & Sons, New York, (1971).

[42] Weiss, A.A., "Least Absolute Error Estimation in the Presence of Serial Correlation", *Journal of Econometrics*, 44 (1990), pp.127–158.

末吉 俊幸
〒278　千葉県野田市山崎 2641
東京理科大学 理工学部 経営工学科

ABSTRACT

LEAST ABSOLUTE VALUE ESTIMATION

Toshiyuki Sueyoshi
Science University of Tokyo

This research first describes historical perspectives of Least Absolute Value (LAV) estimation, which has been long considered as an estimation alternative of conventional Least Squares (LS) regression. Then, this article explores statistical properties regarding the LAV estimation from Goal Programming (GP). Using a small illustrative example, this study presents new theoretical features regarding the LAV estimation. It is hoped that this research effort enhances its applicability to deal with many decisinal issues in reality.